Remarks/Arguments

With reference to the Office Action mailed July 25, 2006, Applicants offer the following remarks and argument.

Confirmation of Election

Applicants confirm their election of claims 1-9.

Status of the Claims

Claims 1-17 were originally presented for examination.

The claims were subject to a restriction requirement. Applicants elected claims 1-9 with traverse.

The Art of Record

The primary reference, United States Patent 6,636,849 to Tang et al. for Data Search Employing Metric Spaces, Multi-grid Indexes, And B-grid Trees describes systems and methods for generating indexes and fast searching of "approximate", "fuzzy", or "homologous" matches for a large quantity of data in a metric space. The data is indexed to generate a search tree taxonomy. Once the index is generated, a query can be provided to report all hits within a certain neighborhood of the query. In an even faster implementation, Tang et al. describe using their disclosed method together with existing approximate sequence comparison algorithms, such as FASTA and BLAST. As described by Tang et al., a local distance of a local metric space is used to generate local search tree branches. This may include homology search for DNA and/or protein sequences, textual or byte-based searches, literature search based on lists of keywords, and vector and matrix based indexing and searching.

However, as will be described below, Tange et al. do not disclose Applicants' claimed

The Office Action of July 25, 2006

35 USC §101 Rejection

Claim 1 was rejected as being directed to non-statutory subject matter. It was suggested that "the cluster estimation subsystem selects said patches depending on said inter-patch confidence values to represent clusters of said document-keyword vectors." This amendment has been made.

Art Rejections

In the Office Action of July 25, 2006, the claims (claims 1-9) were rejected as anticipated by Tang et al.

Objections

In the Office Action of July 25, 2006, the drawings were objected to under 37 CFR § 1.83(a) in that:

(1) Figure 15 shows modified forms of construction in the same view.

(2) Figure 4 does not have reference numbers for the patches.

(3) Figure 5 does not have appropriate reference numbers.

Discussion

The overarching issue presented is that Applicants' amendments impart allowability to the amended claims.

Analysis of Original Claim 1. The following analysis is presented showing the art of

record applied to claim 1 as partially amended (to overcome the 35 USC § 101 rejection only:

A computer system for generating data structures for information retrieval of documents stored in a database, said documents being stored as document-keyword vectors generated from a predetermined keyword list, and said document-keyword vectors forming nodes of a hierarchical structure imposed upon said documents, said computer system comprising:

Tang et al., Abstract

Systems and methods for generating indexes and fast searching of "approximate", "fuzzy", or "homologous" matches for a large quantity of data in a metric space are provided. The data is indexed to generate a search tree taxonomy. Once the index is generated, a query can be provided to report all hits within a certain neighborhood of the query. In an even faster implementation, the invention may be used together with existing approximate sequence comparison algorithms, such as FASTA and BLAST. Here, a local distance of a local metric space is used to generate local search tree branches. Applications of this invention may include homology search for DNA and/or protein sequences, textual or byte-based searches, literature search based on lists of keywords, and vector and matrix based indexing and searching.

a neighborhood patch generation subsystem for generating groups of nodes having similarities as determined using a search structure, said neighborhood patch generation subsystem including a subsystem for generating a hierarchical structure upon said document-keyword vectors and a patch defining subsystem for creating patch relationships among said nodes with respect to a metric distance between nodes; and

Tang, column 4, lines 39-54

In operation, the systems and methods of the present invention first process a data set to create a multigrid tree. The multigrid tree comprises gridpoints (e.g. a data element in the data space comprising of the data set or, stated differently, a collection of adjacent points in the data space). The multigrid tree is calculated using distance functions of a metric space. Associated with each grid point is a radius that defines the neighborhood of the grid point (i.e., a grid). In an illustrative implementation, the multigrid tree comprises a plurality of descending branches that originate from a root grid point. The further the branch from the root grid point, the smaller the radius of the grid points residing on that branch. The multigrid tree may be a B-grid tree that is balanced such that data elements

of the data set are partitioned in equal size grids such that search time is more homogenous for varying search queries.

Tang, Column 11, lines 18-27

The grid concept can be extended one more step such that there <u>*are multiple levels of grids*</u>*. Each grid at a fixed level can be subdivided into smaller grids with smaller radius (i.e. a smaller neighborhood). Those smaller grids become children, and the original grid with its gridpoint is the parent. In this way,* <u>*multiple levels of grids can be linked via parent-child relationships*</u>*. As illustrated in FIG. 5,* <u>*the multilayered grid structure when assembled forms a grid search tree.*</u>

Tang, Column 10, line 61 – column 11, line 5

As shown in FIG. 4, metric space "E" 400 can be divided into many small grids 405, 410, 415, etc., each containing a grid point, 405a, 410a, 415a, etc, respectively. FIG. 4 shows an example of a multigrid in a 2dimensional point set with <u>*L.sub.1 distance*</u> *and a corresponding search performed on the grid. For example, consider a set of points E, all which are located in a 2dimensional area of [1,5][1,4]. The* <u>*L.sub.1 distance*</u> *may be defined as follows, given p.sub.1 =(x.sub.1,y.sub.1), p.sub.2 =(x.sub.2,y.sub.2), d(p.sub.1,p.sub.2)=max(.vertline.x.sub.1 x.sub.2.vertline.,.vertline.y.sub.1 y.sub.2.vertline.).* <u>*Using this calculated distance, an exemplary search may be performed to answer the question: given a query point q=(2.2, 1.8), find out all points p within the area that satisfy d(q,p)<0.3.*</u>

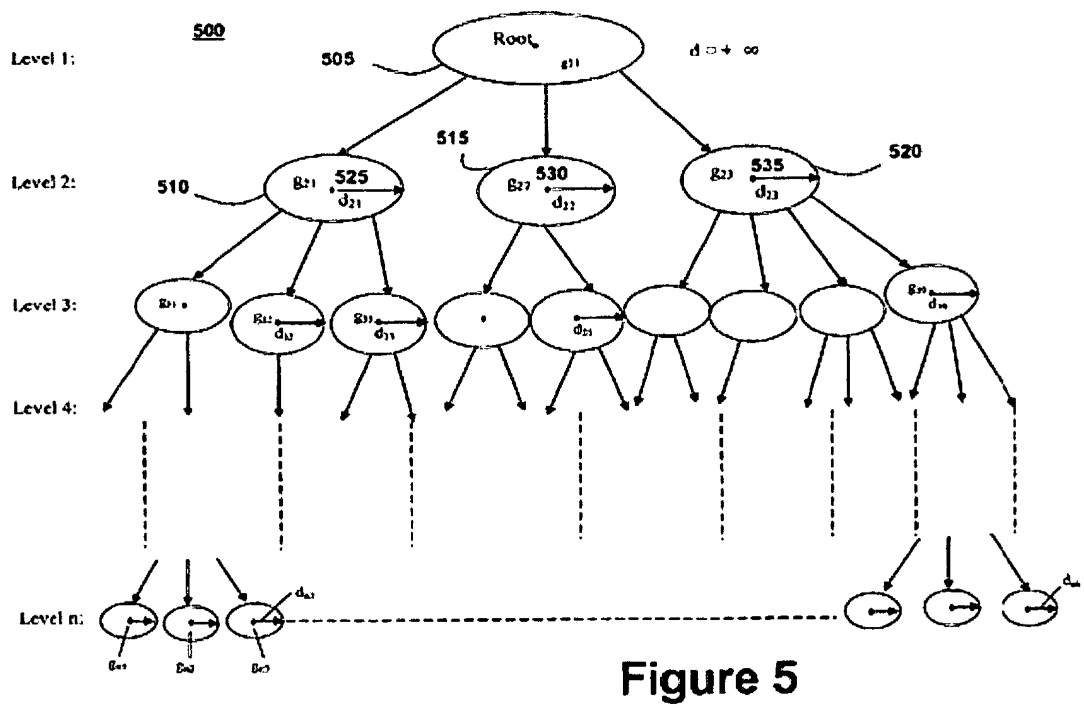

Figure 5

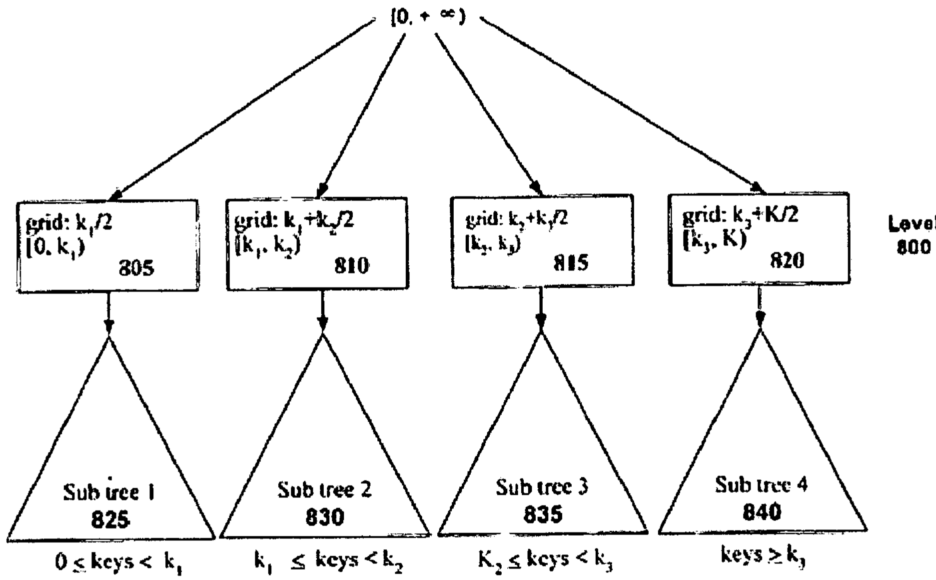

Figure 8A

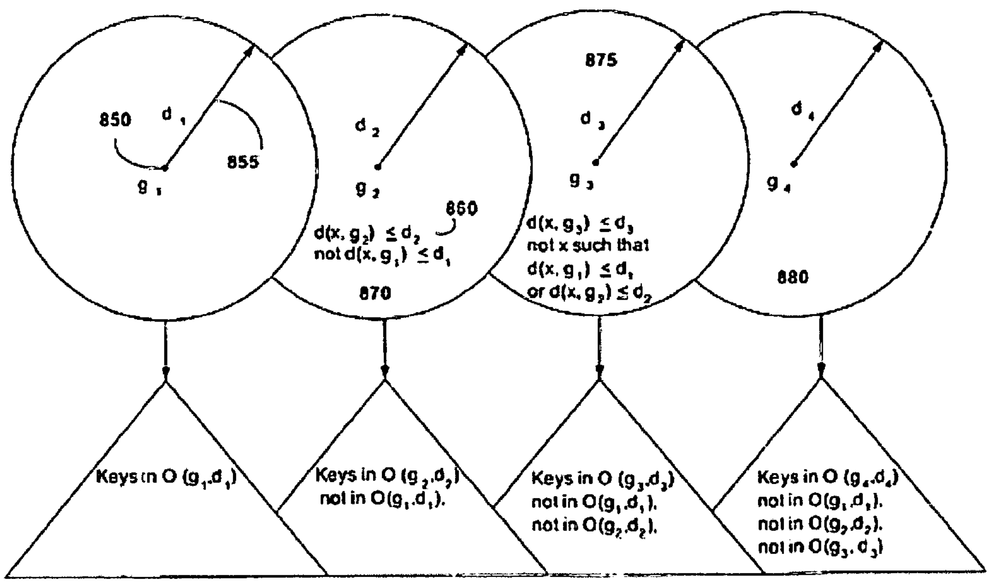

Figure 8B

Tang, Column 4, line 55 to column 5, line 17

For example, for a given query point q, inexact matches to q in a given data set can be found. In mathematical terms, the search aims to find all points p in the data set such that those points p satisfy d(q,p).ltoreq..epsilon., where d(. , .) is the distance function in the metric space, and .epsilon. is the size of interested neighborhood. When a search started, a comparison is performed among all of the grid points of at the first level of the created Bgrid tree. At each level, many subtrees are totally eliminated for further search by applying the triangular inequality rule. For example, suppose the grid points are g.sub.ij, the comparison search for the desired data elements to satisfy the search string q is only to be further carried out within those grids where d(q,g.sub.ij)<.epsilon.+.delta..sub.ij, where .delta..sub.ij is the chosen grid size. The systems and methods of the present invention perform these calculations to produce result set for communication to the participating user.

In an alternative implementation, the systems and methods of the present invention are used to calculate local distances for submitted search queries. For example, for a given query q, the search aims to find most of points p where d(p, q)<.epsilon., whereas the missed points p are likely close to the boundary of d(p,q)=.epsilon.. In this implementation, current search algorithms, such as, BLAST and FASTA are used to create a local multigrid tree (or a local Bgrid tree) having local distances. Employing the same steps above, the local multigrid tree (or local Bgrid) tree is analyzed to find data elements for the submitted search query. Since local distances are used to create the local multigrid tree (or the local Bgrid search tree), the result set will contain most of the desired hits for a submitted search query.

Tang, column 13, line 61 – Column 14, line 16

The process of building the B-grid tree of FIG. 6A is described by the flow diagram of FIG. 5. This method is a slightly modified approach compared with the B-tree definitions described by R. Bayer and E. McCreight, "Organization and Maintenance of Large Ordered Indexes," Acta Informatica. 1:173189 (1970), which is herein incorporated by reference. The B-grid tree of FIG. 6A maintains each sub-tree within its parent grid; whereas in the conventional B-tree definition, the sub-tree is either to the left or right of its parent node. This slight difference makes the B-grid tree concept uniform to all space dimensions in a metric space.

FIG. 8B shows an exemplary B-grid tree of order 4 in a 2dimensioanl metric space. Similar to the B-grid tree of FIG. 8A, each grid is defined by a grid point 850, a radius 855, and some descriptions 860. A shown, there are four grids g.sub.1 (865), g.sub.2 (870), g.sub.3 (875), and g.sub.4 (880). The neighborhoods defined by the grid points and the radii may be overlapping, but as the descriptions indicate, these neighborhoods exist separate and apart. Thus, the grids at the same level have no overlapping regions (points). The descriptions are provided to assign the points in overlapping regions to one of the grids.

Comment: Tang et al do not show a patch generation subsystem for generating groups of nodes, or a subsystem for generating a hierarchal structure, or a patch defining subsystem. These are all claimed elements of the invention. What Tang et al. disclose is the finer and finer sub-division of multi-tiered grids.

a cluster estimation subsystem for generating cluster data of said document-keyword vectors using said similarities of patches <u>wherein the cluster estimation subsystem selects said patches depending on said inner-patch confidence values to represent clusters of said document keyword vectors</u>.

Column 11, lines 6-18

To solve this problem, a set of grid points g.sub.11 =(1.5,1.5), g.sub.12 =(1.5,2.5), . . . with a radius of 0.5 are first chosen for searching. These grid points may be not part of the metric space E. Applying the "triangle inequality" rule of FIG. 3, query "q" 420 may be compared with all of the grid points such that non-relevant neighborhoods are eliminated from the search and to produce a result set containing only relevant grids. The result set shows that search query "q" 420 is a subset of grids: g.sub.11, g.sub.12, g.sub.21, g.sub.22 of metric space E. As such, and as shown in the example, the search is reduced from comparing the search query "q" with all shown neighborhoods (grids) to comparing "q" with only four grids, a significant increase in efficiency.

The grid concept can be extended one more step such that there are multiple levels of grids. Each grid at a fixed level can be subdivided into smaller grids with smaller radius (i.e. a smaller neighborhood). Those smaller grids become children, and the original grid with its grid-point is the parent. In this way, multiple levels of grids can be linked via parent-child relationships. As illustrated in FIG. 5, the multilayered grid structure when assembled forms a grid search tree.

Tang, Column 10, line 61 – column 11, line 27

As shown in FIG. 4, metric space "E" 400 can be divided into many small grids 405, 410, 415, etc., each containing a grid point, 405a, 410a, 415a, etc, respectively. FIG. 4 shows an example of a multi-grid in a 2dimensional point set with L.sub.1 distance and a corresponding search performed on the grid. For example, consider a set of points E, all which are located in a 2-dimensional area of [1,5][1,4]. The L.sub.1 distance may be defined as follows, given p.sub.1 =(x.sub.1,y.sub.1), p.sub.2 =(x.sub.2,y.sub.2), d(p.sub.1,p.sub.2)=max(.vertline.x.sub.1 x.sub.2.vertline.,.vertline.y.sub.1 y.sub.2.vertline.). Using this calculated distance, an exemplary search may be performed to answer the question: given a query point q=(2.2, 1.8), find out all points p within the area that satisfy d(q,p)<0.3.

To solve this problem, a set of grid points g.sub.11 =(1.5,1.5), g.sub.12 =(1.5,2.5), . . . with a radius of 0.5 are first chosen for searching. These grid points may be not part of the metric space E. Applying the "triangle inequality" rule of FIG. 3, query "q" 420 may be compared with all of the grid points such that non-relevant neighborhoods are eliminated from the search and to produce a result set containing only relevant grids. The result set shows that search query "q" 420 is a subset of grids: g.sub.11, g.sub.12, g.sub.21, g.sub.22 of metric space E. As such, and as shown in the example, the search is reduced from comparing the search query "q" with all shown neighborhoods (grids) to comparing "q" with only four grids, a significant increase in efficiency.

The grid concept can be extended one more step such that there are multiple levels of grids. Each grid at a fixed level can be subdivided into smaller grids with smaller radius (i.e. a smaller neighborhood). Those smaller grids become children, and the original grid

with its grid-point is the parent. In this way, multiple levels of grids can be linked via parent-child relationships. As illustrated in FIG. 5, the multilayered grid structure when assembled forms a grid search tree.

Tang, column 13, line 61 – Column 14, line 16

The process of building the B-grid tree of FIG. 6A is described by the flow diagram of FIG. 5. This method is a slightly modified approach compared with the B-tree definitions described by R. Bayer and E. McCreight, "Organization and Maintenance of Large Ordered Indexes," Acta Informatica. 1:173189 (1970), which is herein incorporated by reference. The B-grid tree of FIG. 6A maintains each sub-tree within its parent grid; whereas in the conventional B-tree definition, the sub-tree is either to the left or right of its parent node. This slight difference makes the B-grid tree concept uniform to all space dimensions in a metric space.

FIG. 8B shows an exemplary B-grid tree of order 4 in a 2dimensioanl metric space. Similar to the B-grid tree of FIG. 8A, each grid is defined by a grid point 850, a radius 855, and some descriptions 860. A shown, there are four grids g.sub.1 (865), g.sub.2 (870), g.sub.3 (875), and g.sub.4 (880). The neighborhoods defined by the grid points and the radii may be overlapping, but as the descriptions indicate, these neighborhoods exist separate and apart. Thus, the grids at the same level have no overlapping regions (points). The descriptions are provided to assign the points in overlapping regions to one of the grids.

Comment: The cited passages describe subdividing grid points into ever smaller, more granular sets

to produce a result set containing only relevant grids. The result set shows that search query "q" 420 is a subset of grids: g.sub.11, g.sub.12, g.sub.21, g.sub.22 of metric space E. As such, and as shown in the example, the search is reduced from comparing the search query "q" with all shown neighborhoods (grids) to comparing "q" with only four grids, a significant increase in efficiency.

thereby carrying the "grid concept" one step lower. This is clearly contrary to Applicants' claimed "cluster estimation subsystem for generating cluster data of said document-keyword vectors using said similarities of patches wherein the cluster estimation subsystem selects said patches depending on said inner-patch confidence values to represent clusters of said document keyword vectors."

Analysis of Amended Claim 1.

Claim 1 as amended is representative of the amended independent claims. Amended claim 1 is reproduced below:

Claim 1. (Currently Amended) A computer system for generating data structures for information retrieval of documents stored in a database, said documents being stored as document-keyword vectors generated from a predetermined keyword list, and said document-keyword vectors forming nodes of a hierarchical structure imposed upon said documents, said computer system comprising:

a neighborhood patch generation subsystem for generating groups of nodes having similarities as determined using a search structure, said neighborhood patch generation subsystem including a subsystem for generating a hierarchical structure upon said document-keyword vectors and a patch defining subsystem for creating patch relationships among said nodes with respect to a metric distance between nodes;

a query vector generation subsystem accepting search conditions and query keywords, generating a corresponding query vector, and storing the generated query vector;

a confidence determination subsystem for computing intra-patch confidence values between patches and interpath confidence values; and

a cluster estimation subsystem for generating cluster data of said document-keyword vectors using said similarities of patches wherein the cluster estimation subsystem selects said patches depending on said inner-patch confidence values to represent clusters of said document keyword vectors, estimates the sizes of said patches, and generates cluster data of document keyword vectors using similarities of the patches.

Of special note are the limitations added by the present amendment. These are:

*** a query vector generation subsystem accepting search conditions and query keywords, generating a corresponding query vector, and storing the generated query vector;

*** a confidence determination subsystem for computing intra-patch confidence values between patches and interpath confidence values; and

*** a cluster estimation subsystem for generating cluster data of said document-keyword vectors using said similarities of patches wherein the cluster estimation subsystem

**** selects said patches depending on said inner-patch confidence values to represent clusters of said document keyword vectors,

**** estimates the sizes of said patches, and

**** generates cluster data of document keyword vectors using similarities of the patches.

"Selects Said Patches Depending On Said Inner-Patch Confidence Values To Represent Clusters Of Said Document Keyword Vectors" The claim limitation recites "confidence values", especially "inner patch confidence values." Confidence values are described as, for example "That is, the confidence is expressed as the proportion of elements forming Ci that also contribute to the formation of Cj. If the confidence value is small, the candidate Cj has little or no impact upon Ci; on the other hand, if the proportion is large, Cj is strongly related to Ci, possibly even subsuming it." at [0116]. There is no disclosure of "confidence" or "confidence values" in Tang et al.

"Generates cluster data of document keyword vectors using similarities of the patches". The claim limitation recites "similarities of the patches." Tang et al. describes sequence similarity, which is one dimensional. Applicants claimed (by method of determining) and disclosed similarities are two dimensional.

Thus applicants' claimed invention, defined by methods and measures of obtaining and using inner patch confidence values and two dimensional similarities, is allowable over Tang et al.

Conclusion

Based on the above discussion, it is respectfully submitted that the pending claims describe an invention that is properly allowable to the Applicants.

If any issues remain unresolved despite the present amendment, the Examiner is requested to telephone Applicants' Attorney at the telephone number shown below to arrange for a telephonic interview before issuing another Office Action.

Applicants would like to take this opportunity to thank the Examiner for a thorough and competent examination and for courtesies extended to Applicants' Attorney.

Respectfully Submitted

Certificate of Mailing	
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being deposited with the United States Postal Service on the date shown below with sufficient postage as Certified Priority Mail (Certified, Return Receipt Requested, Certified Label Number **7006-0100-0007-0960-6642**) in an envelope addressed to the Commissioner for Patents, Alexandria Virginia, 22313 Date of deposit: January 25, 2007 Person mailing paper: Richard M. Goldman Signature: ____________	____________ Richard M. Goldman, Reg. # 25,585 371 Elan Village Lane, Suite 208 San Jose, CA 95134 Voice: 408-324-0716 Fax: 408-324-0672 E-mail: goldmanptn@aol.com